

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-015842

(43)Date of publication of application : 22.01.1999

(51)Int.Cl. G06F 17/30

(21)Application number : 09-167070

(71)Applicant : MITSUBISHI ELECTRIC CORP

(22)Date of filing : 24.06.1997

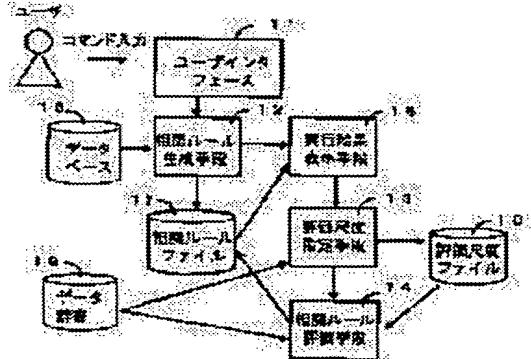
(72)Inventor : OBATA YASUSHI  
YASUDA SATOSHI

## (54) DATA MINING DEVICE

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To obtain a data mining device enabling a user to easily execute operation for retrieving an effective correlation rule useful for user's own purpose out of many correlation rules extracted as the result of data mining by setting up the evaluation scales of correlation rules different in each user's using purpose.

**SOLUTION:** A correlation rule generation means 12 generates and outputs a correlation rule file 17 based on a data base 16. An evaluation scale specification means 13 outputs an evaluation scale file 18 based on an evaluation scale specified by the user. A correlation rule evaluation means 14 calculates an evaluation value based on the file 18 and updates information related to the values of the correlation rules in the file 17. An execution result display means 15 displays the value information of reevaluated correlation rule based on the file 17, executes the rearrangement of display order and the limitation processing of correlation rules to be displayed and then displays the correlation rules.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's

[decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (JP)

## (12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-15842

(43)公開日 平成11年(1999)1月22日

(51) Int.Cl.<sup>6</sup>  
G 0 6 F 17/30

識別記号

F I  
G 0 6 F 15/401  
15/4033 2 0 Z  
3 8 0 D

審査請求 未請求 請求項の数8 OL (全9頁)

(21)出願番号 特願平9-167070

(22)出願日 平成9年(1997)6月24日

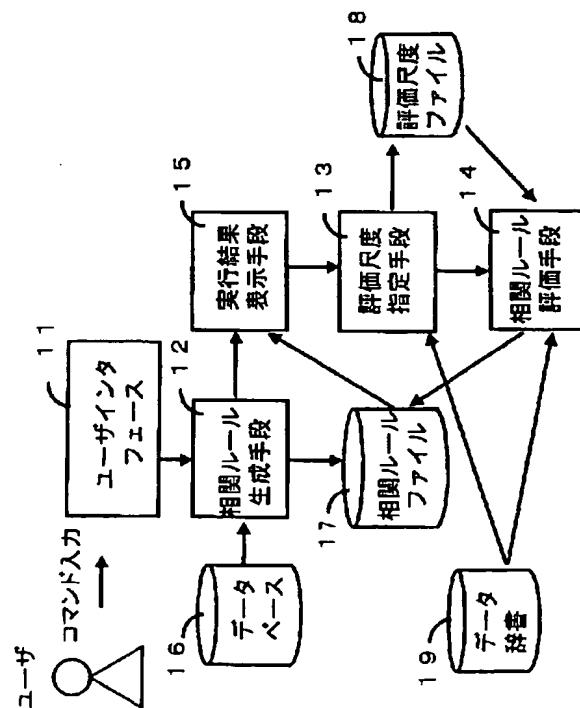
(71)出願人 000006013  
三菱電機株式会社  
東京都千代田区丸の内二丁目2番3号  
(72)発明者 小幡 康  
東京都千代田区丸の内二丁目2番3号 三菱電機株式会社内  
(72)発明者 安田 智  
東京都千代田区丸の内二丁目2番3号 三菱電機株式会社内  
(74)代理人 弁理士 宮田 金雄 (外2名)

(54)【発明の名称】 データマイニング装置

## (57)【要約】

【課題】 ユーザの使用目的毎に異なる相関ルールの評価尺度を設定できるようにすることにより、データマイニングの結果抽出された大量の相関ルールの中から、ユーザが自分の目的に役立つ有効な相関ルールを見つけることを目的とする。

【解決手段】 相関ルール生成手段12はデータベース16を基に相関ルールファイル17を生成し出力する。評価尺度指定手段13は、ユーザが指定した評価尺度に基づき評価尺度ファイル18を出力する。相関ルール評価手段14は、評価尺度ファイル18の基づいて評価値を算出し、相関ルールファイル17中の相関ルールの価値に関する情報を更新する。実行結果表示手段15は、相関ルールファイル17に基づいて、再評価された相関ルールの価値情報を表示すると共に、表示順序の並び替えや、表示する相関ルールの限定処理を行った上で相関ルールを表示する。



## 【特許請求の範囲】

【請求項1】 データベース中のデータ属性間に存在する相関ルールの相関分析を行うデータマイニング装置において：データベースからデータを受け、データ項目間の相関ルールを生成する相関ルール生成手段と、ユーザが、その使用目的毎に異なる相関ルールの評価尺度を指定できる評価尺度指定手段と、前記評価尺度指定手段においてユーザが指定した評価尺度を用いて前記相関ルール生成手段によって生成された相関ルールの価値を算出する相関ルール評価手段と、前記相関ルール生成手段によって生成された相関ルールを前記相関ルール評価手段が算出した相関ルールの価値に基づいて表示する実行結果表示手段を備えたことを特徴とするデータマイニング装置。

【請求項2】 請求項1記載のデータマイニング装置において：前記評価尺度指定手段は、ユーザによって指定された、相関ルールを利用した場合に発生する損失値または利益値（以下、相関ルール適用時のコスト）と、その相関ルールが成立した場合に発生する利益値または損失値（以下、相関ルール成立時の利益）とにより評価尺度の指定を行うことを特徴とするデータマイニング装置。

【請求項3】 請求項2記載のデータマイニング装置において：前記の相関ルール適用時のコストおよび相関ルール成立時の利益は、データ項目毎に各種項目について値が定義されたデータ辞書中に定義された項目、または定数値、またはその両方を項とする四則演算式を用いて記述されることを特徴とするデータマイニング装置。

【請求項4】 請求項2記載のデータマイニング装置において：前記評価尺度指定手段は、指定した相関ルール適用時のコストおよび相関ルール成立時の利益に関する情報を記憶装置に保存し、後でこれを読み込んで再利用する機能を持つことを特徴とするデータマイニング装置。

【請求項5】 請求項1記載のデータマイニング装置において：前記相関ルール評価手段は、前記評価尺度指定手段によって指定された相関ルール適用時のコストおよび相関ルール成立時の利益、前記データベース中の相関ルールの精度（正解率）を示す確信度、並びに前記データベース中でその相関ルールが成立しているレコード数の全レコード件数に対する割合を表す支持度に基づいて評価値を算出することを特徴とするデータマイニング装置。

【請求項6】 請求項5記載のデータマイニング装置において：前記相関ルール評価手段は、前記相関ルール適用時のコストおよび相関ルール成立時の利益を記述した四則演算式を計算する際に、前記データ辞書中に定義された項目の値を参照することを特徴とするデータマイニング装置。

【請求項7】 請求項1記載のデータマイニング装置に

おいて：前記実行結果表示手段は、演算の結果得られた相関ルールのうち、前記相関ルール評価手段が算出した評価値がユーザによって指定された基準値以上である相関ルールのみを表示する機能を持つことを特徴とするデータマイニング装置。

【請求項8】 請求項1に記載のデータマイニング装置において：前記実行結果表示手段は、演算の結果得られた相関ルールを、前記相関ルール評価手段が算出した評価値に基づき、昇順または降順の並べ替え処理を行った上で表示する機能を持つことを特徴とするデータマイニング装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 この発明は、データベース中の大量データを分析し、データの属性間に存在する相関ルールを抽出するデータマイニング装置に関するものである。

## 【0002】

【従来の技術】 データマイニングとは、大規模データベース中の大量データから、そのデータに含まれるデータ項目間に存在する規則性や因果関係を発掘することをいうが、その代表的なものとして、相関分析がある。これは、相関ルールと呼ばれる相関ルールを抽出するものである。相関ルールとは、データ項目間に存在する相関関係を相関ルールの形で表現したものであり、具体例として、たとえば「データ項目（部分集合）Aとデータ項目（部分集合）Bが同一トランザクションに存在する時には、高頻度でデータ項目（部分集合）Cも同時に存在する」といったような規則性（相関関係）を「A, B → C」などと表現したものを言う。代表的な相関分析の例としてバスケット分析と呼ばれるものがある。これは、小売業において、客が同時にバスケット（=買い物かご）に入れる商品の相関関係を求めるもので、同時購買分析とも呼ばれる。バスケット分析では、蓄積されたレシートデータを用いた相関分析により、たとえば「パン → 牛乳」（パンを買う客は、同時に牛乳も買う）という相関ルールが求められる。

【0003】 データマイニングにおける相関分析の基本的な処理は、相関ルールの候補の生成とその検証である。つまり、相関ルールの候補を全てのデータ項目の組み合わせから生成し、この候補一つ一つについて、実際に成り立っているかどうかをデータベース検索による数え上げによって調べるという方法である。しかし、これでは効率が悪いので、従来のデータマイニングでは、有効な相関ルールを効率よく発見するため、この処理の中で支持度と確信度という足切り基準によって、相関相関ルールの候補の数をしぶる枝刈りを行いながら処理を行っている。

【0004】 支持度とは、その相関ルールの一般性を表わす尺度であり、確信度とは、その相関ルールの正確さ

を表わす尺度である。相関ルールは一般に、この支持度と確信度と呼ばれる値を伴って、「A→B」という形の論理式で表わされる（但し、AとBはそれぞれ空でないデータ項目の集合であり、交わらないものとする）。この場合、支持度は、全レコードのうち部分集合Aと部分集合Bの両方に属する要素の集合「A ∪ B」を含むレコードの占める割合で表わされ、確信度は、Aを含むレコードのうち、Bも同時に含むレコードの割合で表わされる。たとえば、前述の「パン→牛乳」の例で言えば、全レシート件数（レコード件数）のうち、パンを購入した客の割合が20%で、パンと牛乳両方を購入した客の割合が12%だとすると、相関ルール「パン→牛乳」の支持度は12%、確信度は60%（=12%/20%）ということになる。

【0005】従来のデータマイニング装置では、相関ルールを生成する時に、支持度と確信度の下限値をそれぞれ設定し、この下限値をともに上回るすべての相関ルールを抽出する。相関ルールを抽出する方法については、例えば特開平8-263346号公報、あるいは特開平8-287106号公報に詳しく述べられているようには、まず支持度が下限値を上回る相関ルールの候補を生成する。この相関ルールの候補生成については、特に特開平8-263346号公報に詳しく述べられている。その後にその相関ルールの候補の確信度を調べて下限値を上回る候補を採用して、最終的な相関ルールとして出力する。すなわち、この方法で得られる相関ルールは支持度と確信度の2つの指標のみによって抽出されたものであり、その他の評価、例えばユーザにとって興味のある売り上げへの寄与や、ユーザの意思等の評価は一切加わっていない。

【0006】このようなデータマイニングの結果得られる相関ルールは、大抵の場合、大量である。しかもそのうちのほとんどは、ユーザが求めているものではなく、既に陳腐化したものであったり、意味のない相関であったりする。ユーザは、この無数の相関ルールの中から、それぞれ自分の目的に役立つ有効な相関ルールを見つける作業を行わなければならない。

【0007】この問題を解決し、ユーザにとって有効な相関ルールだけを抽出するためには、相関ルールの有効度を評価する尺度が必要となる。たとえば、特開平8-77010号公報では、相関ルールが対象とするレコード数で表現されるカバー率（前述の支持度に相当）と、相関ルールの正解率で表現されるヒット率（前述の確信度に相当）から相関ルールの評価尺度を算出している。

【0008】また、電気情報通信学会信学技報、1995-05、41頁-48頁の福田剛志、森下真一両氏による「相関ルールの可視化について」なる文献では、統計的検定を用いて支持度、確信度の評価を行うことにより、「面白くない相関ルール」すなわち有効でない相関ルールをふるいにかけるという方法が解説されている。

### 【0009】

【発明が解決しようとする課題】従来のデータマイニング装置では、相関ルールの有効性の評価尺度を、確信度と支持度に基づいて定めている。つまり、一般性が高く（支持度が高く）、精度の高い（確信度の高い）相関ルールが有効な相関ルールとして評価される。このような評価尺度は、データの特徴を精度よく表現することを目的としたデータマイニングにおいては、相関ルールの価値を評価するひとつの基準として有効である。

【0010】しかし、データマイニングは上記のような目的のみに使用されるものではなく、むしろデータマイニングの結果得られた相関ルールを意思決定支援、戦略立案等、何かに役立てたいという要求のもとに行われるのが普通である。データマイニングによって得られた相関ルールを何かの目的に応用しようとする時、たとえば、前述のバスケット分析の結果得られた相関ルールを販売促進による売上げ増のための戦略立案に役立てたいと考えた場合、支持度、確信度によって高く評価された相関ルールが、必ずしもユーザの目的（この場合は売上げ増）に対して有効性の高い相関ルールであるとは限らない。この場合、ユーザの目的に対して有効性の高い相関ルールとは、たとえば大幅な売上げ増を見込める相関ルールである。

【0011】このように、一般的に相関ルールの価値というものは、その相関ルールをどのように使用するか、というユーザの目的によって異なるものであるが、従来のような支持度、確信度による画一的な評価尺度では、必ずしもユーザの目的に関して適正に評価できないという問題点がある。すなわち、従来の技術では、支持度、確信度のみで相関ルールの価値を判断しているので、

「その相関ルールを適用した場合、どの位の利益が見込めるか」ということを知りたい、「利益の大きな相関ルールを強調して表示したい」という目的でデータマイニングを行なっている場合、支持度、確信度は利益の見込みとは無関係であるので、この目的に沿って相関ルールを評価できないという問題点がある。

【0012】この発明は上記のような問題点を解消するためになされたもので、ユーザの使用目的毎に異なる相関ルールの評価尺度を適用することのできるデータマイニング装置を得ることを目的としており、これにより、データマイニングの結果抽出された大量の相関ルールの中から、ユーザが自分の目的に役立つ有効な相関ルールを見つける作業を容易に行えるデータマイニング装置を得ることを目的としている。

### 【0013】

【課題を解決するための手段】本発明の第1の発明のデータマイニング装置は、データベースからデータを受け、データ項目間の相関ルールを生成する相関ルール生成手段と、ユーザが、その使用目的毎に異なる相関ルールの評価尺度を指定できる評価尺度指定手段と、上記評

価尺度指定手段においてユーザが指定した評価尺度を用いて上記相関ルール生成手段によって生成された相関ルールの価値を算出する相関ルール評価手段と、上記相関ルール生成手段によって生成された相関ルールを上記相関ルール評価手段が算出した相関ルールの価値に基づいて表示する実行結果表示手段を備えたるよう構成される。

【0014】また、本発明の第2の発明のデータマイニング装置における評価尺度指定手段は、ユーザが相関ルールを利用した場合に発生する損失値または利益値（以下、相関ルール適用時のコスト）と、その相関ルールが成立した場合に発生する利益値または損失値（以下、相関ルール成立時の利益）を指定することにより評価尺度の指定を行うよう構成される。

【0015】また、本発明の第3の発明のデータマイニング装置における評価尺度指定手段において、相関ルール適用時のコストおよび相関ルール成立時の利益は、データ項目毎に各種項目について値が定義されたデータ辞書中に定義された項目、または定数値、またはその両方を項とする四則演算式を用いて記述されるよう構成される。

【0016】また、本発明の第4の発明のデータマイニング装置における評価尺度指定手段は、指定した相関ルール適用時のコストおよび相関ルール成立時の利益に関する情報を記憶装置に保存し、後でこれを読み込んで再利用する機能を持つよう構成される。

【0017】また、本発明の第5の発明のデータマイニング装置における相関ルール評価手段は、前記評価尺度指定手段によって指定された相関ルール適用時のコストおよび相関ルール成立時の利益、前記データベース中の相関ルールの精度（正解率）を示す確信度、並びに前記データベース中でその相関ルールが成立しているレコード数の全レコード件数に対する割合を表す支持度に基づいて評価値を算出するよう構成される。

【0018】また、本発明の第6の発明のデータマイニング装置における相関ルール評価手段は、相関ルール適用時のコストおよび相関ルール成立時の利益を記述した四則演算式を計算する際に、データ辞書中に定義された項目の値を参照するよう構成される。

【0019】また、本発明の第7の発明のデータマイニング装置における実行結果表示手段は、演算の結果得られた相関ルールのうち、相関ルール評価手段が算出した評価値がユーザによって指定された基準値以上である相関ルールのみを表示する機能を持つよう構成される。

【0020】また、本発明の第8の発明のデータマイニング装置における実行結果表示手段は、演算の結果得られた相関ルールを、相関ルール評価手段が算出した評価値に基づき、昇順または降順の並べ替え処理を行った上で表示する機能を持つよう構成される。

【0021】

#### 【発明の実施の形態】

実施の形態1. 以下、この発明の一実施の形態を図1を用いて説明する。図1において、12はデータベース中のデータ項目間に存在する相関ルールを生成する相関ルール生成手段、15は、上記相関ルール生成手段が生成した相関ルールの表示を行い、さらにこの表示順序の並べ替えや表示内容の限定（絞り込み）等の操作を行う実行結果表示手段、13は上記相関ルール生成手段が生成した相関ルールの価値の評価尺度をユーザが指定するための評価尺度指定手段、14は上記評価尺度指定手段13において指定された評価尺度に基づいて上記相関ルール生成手段12が生成した相関ルールの価値を算出する相関ルール評価手段である。11は、上記相関ルール生成手段12に対して、ユーザが情報を与え、上記相関ルール生成手段12を起動するユーザインタフェースである。16は、本発明のデータマイニング装置で処理すべきデータを保存するデータベースである。17は相関ルール生成手段12で生成された相関ルールに関する相関ルールファイルである。18は、評価尺度指定手段13によって生成された評価尺度ファイル18である。評価尺度ファイル18は記憶装置に記憶しておき、必要に応じて必要することができる。19は評価尺度指定手段13によって評価尺度を指定する際に参照されるデータ辞書である。

【0022】次に、本発明のデータマイニング装置の動作について説明する。ユーザはコマンド等によって、ユーザインタフェース11を介して相関ルール生成手段12を起動し、データベース16に保存されたデータを相関ルール生成手段12に入力する。相関ルール生成手段12はデータベース16からのデータを処理して相関ルールファイル17を出力する。たとえば、上述の例によれば、データベース16には、データ項目（パン）Aとデータ項目（牛乳）Bが保存されており、ユーザからのコマンドによって、「パン→牛乳」（パンを買う客は、同時に牛乳も買うという相関ルール）という相関ルールが相関ルール生成手段12から相関ルールファイル17として出力される。この相関ルールファイル17は実行結果表示手段15に入力される。実行結果表示手段15は、相関ルールファイル17中の相関ルールを表示する。評価尺度指定手段13は評価尺度ファイル18を作成するものであり、たとえば、図3に示すようなディスプレイ画面を用いて、ユーザに相関ルールの評価尺度を入力させ、ユーザが入力したその評価尺度を評価尺度ファイル18に出力する。この評価尺度指定手段13は実行結果表示手段15によって起動される。相関ルール評価手段14は、相関ルールファイル17と評価尺度ファイル18とデータ辞書19を基に、相関ルールファイル17に格納された個々の相関ルールに対して相関ルールの価値を算出し、相関ルールファイル17中の個々の相関ルールについて、相関ルールの価値に関する情報を更

新する。実行結果表示手段15は、この相関ルールファイル17に格納された相関ルールを表示する。

【0023】次に、図2のフローチャートを用いて、図1の各手段における処理の流れを説明する。まず、ステップ21において、相関ルール生成手段12は、データベース16からのデータに基づいて、相関ルールファイル17を生成する。この相関ルール生成手段12は公知の手段、たとえば、従来の技術として述べた特開平8-263346号公報、あるいは特開平8-287106号公報を用いるものとする。この場合、このアルゴリズムにおいて使用される「支持度」と「確信度」の2つの情報は、ユーザインタフェース11を介してユーザから与えられる。

【0024】次に、ステップ22において、実行結果表示手段15は、相関ルール生成手段12が出力した相関ルールファイル17を入力し、相関ルールファイル17に格納された相関ルールを実行結果として表示する。ここで、相関ルール生成手段12が生成した相関ルールファイル17は、指定された支持度、確信度以上の相関ルールを全て格納している。しかし、この段階では、まだ相関ルールの評価尺度としては、その相関ルールの支持度および確信度のみが計算されており、各相関ルールについて利益等の他の評価基準はこの段階では計算されていない。必要ならば、実行結果表示手段15は、この支持度、確信度の大きさに基づいて表示順の並び替えや表示する相関ルールの絞り込みをすることもできる。

【0025】次にステップ23において、ユーザは、評価尺度指定手段13を用いて、評価尺度の設定を行う。この評価尺度の設定は、以下に述べるように、相関ルール適用時のコストと、相関ルール成立時の利益を指定することにより行う。相関ルール適用時のコストおよび相関ルール成立時の利益は、それぞれ数式の形で指定する。数式は、相関ルールの左辺（条件部）と右辺（結論部）それぞれに現れるデータ項目についての各種定義情報が格納されたデータ辞書19中の項目が定数値を項とする四則演算式の形式をとる。

【0026】図3に、この数式で評価尺度を指定するための評価尺度指定画面を示す。前記相関ルール適用時のコストは、相関ルール適用時コストエディットコントロール32に入力し、また前記相関ルール成立時の利益は、相関ルール成立時利益エディットコントロール33に、それぞれ数式の形式で入力する。数式は、「右辺<項目名>」または「左辺<項目名>」という2パターンの変数と、定数を項とする四則演算式でそれぞれエディットコントロール32およびエディットコントロール33で指定する。ここで、「右辺<項目名>」は、相関ルールの右辺（=結論部）に現れるデータ項目について、データ辞書中で、カッコ<>内に指定された項目に対して定義されている値、「左辺<項目名>」は、相関ルールの左辺（=条件部）に現れるデータ項目について、データ辞

書中で、カッコ<>内に指定された項目に対して定義されている値、という意味である。

【0027】簡単な例をあげて説明する。図4は、データ辞書の一例を示すイメージである。この例では、データ辞書には、パン、バター、ミルクなどの各データ項目毎に、「定価」「売値」「仕入値」などの項目についての値が定義されている。ここで、たとえば、「パン、バター→ミルク（パン、バターを購入する人はミルクを購入することを意味する）」という相関ルールがあったとする。この相関ルールにおいて、たとえば「右辺<売値> - 右辺<仕入値>」という数式は、相関ルールの右辺のデータ項目が「ミルク」であるので、この数式は「ミルク<売値> - ミルク<仕入値>」となる。従って、「ミルク」というデータ項目について、データ辞書中で定義されている「売値」の値（この例では120）から、同じく「ミルク」についてデータ辞書中で定義されている「仕入値」の値（この例では80）をひいた結果、という意味になる。さらに、詳細に記述すると、「ミルク<売値> - ミルク<仕入値>」 = 120 - 80 = 40となる。

【0028】同様に、「-（左辺<売値> × 0.2 + 100）」という式は、上記相関ルールの左辺に現れているデータ項目について、データ辞書中で定義されている「売値」の値を0.2倍し、100を加えた結果のマイナス値、という意味になる。この相関ルールの左辺は複数のデータ項目を含んでいるが、このような場合、「左辺<売値>」は、左辺のデータ項目それぞれの売値を足しあわせたもの、として考える。つまりこの例では、データ項目「パン」の「売値」とデータ項目「バター」の「売値」を足したもののが「左辺<売値>」である。

【0029】エディットコントロール32および33への数式の入力はキーボードから直接入力するが、ボタン34、ボタン35をクリックすると、それぞれ、図5に示すように、データ辞書に定義されている項目が一覧表示されたリストボックス40が現れるので、その中に表示された項目をクリックすることにより選択することもできる。

【0030】たとえば、エディットコントロール32をクリックしてフォーカスをあらかじめエディットコントロール32に設定しておき、コントロールボタン35をクリックすると、図5のようにリストボックス40があらわれる。この中からたとえば「売値」をクリックすると、「左辺<売値>」という文字がエディットコントロール32に転記される。エディットコントロール32をクリックしてフォーカスをあらかじめエディットコントロール32に設定しておき、コントロールボタン34で同様の操作を行った場合には、「右辺<売値>」という文字がエディットコントロール32に転記される。上記の例ではフォーカスがエディットコントロール32に設定された例を示したが、この転記については、エディットコントロール32、または33のどちらがフォーカス

を持っているかによって決まる。フォーカスをあらかじめエディットコントロール33に設定しておくと、転記はフォーカスの当たっているエディットコントロール33に対して行われる。

【0031】ボタン38は、評価尺度指定画面で指定した評価尺度に名前を付けて保存しておくためのボタンである。エディットコントロール31に評価尺度の名前を入力し、ボタン38をクリックすると、指定した評価尺度を保存しておくこともできる。このようにして保存した評価尺度は、後で読み込んで再利用することが可能である。ボタン39をクリックすると、現在保存されている評価尺度の名前の一覧が表示されるので、この中から使用する評価尺度を選択する。ボタン37をクリックすると評価尺度の指定は中止される。ボタン36をクリックするとエディットコントロール32、33で指定された数式で評価尺度が確定され、評価尺度の指定を終了する。

【0032】再び図2に戻る。ステップ24では、相関ルール評価手段14は、ステップ23でユーザが指定した評価尺度を用いて相関ルールの価値の算出を行い、相関ルールファイル17中の相関ルールの価値に関する情報を更新する。このステップ24における相関ルールの価値の算出は、次のように行う。

相関ルールの価値=相関ルール成立時の利益×（相関ルールの支持度×相関ルールの確信度）+相関ルール適用時のコスト×相関ルールの支持度

この式で相関ルール成立時の利益とは、図3のエディットコントロール33で指定された数式によって、相関ルール毎に求められる値であり、相関ルール適用時のコストとは、同じく図3のエディットコントロール32で指定された数式によって、相関ルール毎に求められる値である。たとえば、「右辺<売値>−右辺<仕入値>」が32で指定されているとします。相関ルール「パン、バター→ミルク」では、「ミルクの売値−ミルクの仕入値」が、相関ルール「アイスクリーム→ジュース」では、「ジュースの売値−ジュースの仕入値」という様に、相関相関ルール毎に計算されます。この数式を評価する（計算する）際、式中で使用されている変数、たとえば、上述の計算では「ミルクの売値と仕入値」、「ジュースの売値と仕入値」については、データ辞書中で定義されている値を参照して使用する。

【0033】従来技術のところでも説明したように、相関ルールの支持度は、その相関ルールの一般性を表す数値であり、相関ルールを利用する際に、支持度が大きいほど適用できる場面が多いと考えることができる。また、確信度はその相関ルールが成立する確率であり、その相関ルールを利用した場合に、ユーザが期待した結果（つまり利益）を得られる期待度と考えることができる。この式は、このような前提に基づいて定めたものである。

【0034】最後にステップ25において、実行結果表示手段15は、ステップ24において相関ルールの評価値が更新された相関ルールファイル17を取り込み、ユーザの必要性に応じ、ステップ24で算出された相関ルールの評価値に基づいた相関ルールを並べ替え表示（昇順／降順）し、またはある基準以上（または以下）の評価値をもつ相関ルールのみを表示する表示内容の絞り込み等の操作を行う。たとえば、ユーザの目的が「大きな利益が見込める相関ルールを抽出したい」というものであった場合、その利益の見込みを評価基準として全ての相関ルールについて計算すれば、その値がある一定のしきい値以上の相関ルールのみを表示したり、値の大きい順に相関ルールを並べて表示したりすることができます。これによって、データマイニング装置は、利益の大きな相関ルールを強調して表示する等、ユーザの目的に応じた有効な相関ルールの洗い出し作業を容易に支援することが可能である。

【0035】ここで、上記について、簡単な適用例を想定して説明する。ある店主が前述のバスケット分析を行い、その結果得られた相関ルールを利用してダイレクトメールマーケティングを実施することによって、業績の向上を考えているものとする。その店主は、データマイニングに用いるデータ入手し、そのデータを用いてバスケット分析を行った結果、大量の相関ルールを抽出した。さて、この大量の相関ルールの中から今回ダイレクトメールマーケティングに役立てることのできる相関ルールを探し出すケースを考える。

【0036】ダイレクトメールマーケティングは、顧客にダイレクトメールを送り付けることにより、それを見て商品を買ってくれるお客様によって売上げ向上を図るものである。当然、ダイレクトメールに反応して商品を購入してくれる客と、そうでない客がいるわけだが、ダイレクトメールを出すのにもコストがかかるので、もし、ダイレクトメールに反応してくれる確率の高い客層というものがわかれれば、効率のよいダイレクトメールマーケティングを展開することができる。たとえば、「商品A→商品B」という相関ルールによって「商品Aを買うお客様は同時に商品Bを購入する」という傾向を知ることができれば、商品Aを買ったお客様に対し、商品Bに関する広告をダイレクトメールで送ることによって、潜在的購買層に対し、商品Bの購買欲を刺激する、といった販売促進戦略を立てることが可能である。

【0037】まず、この例では、ダイレクトメールを一通出すのに、100円のコストがかかるとする。ダイレクトメールに反応し、実際に商品を買ってくれた場合の利益は、相関ルールの右辺にくる商品が売れた場合の利益である。データ辞書には先の例の図4と同様、「定価」「売値」「仕入値」といった項目が定義されているとする。図2のステップ23において、ユーザが3の評価尺度指定画面で指定すべき評価尺度は、

相関ルール適用時のコスト：-100

相関ルール成立時の利益：右辺<売値>-右辺<仕入値>となる。

【0038】このような評価尺度で相関ルールの評価値を算出すると、その相関ルールが示す内容を利用してダイレクトメールを出した場合（たとえば、相関ルール「商品A→商品B」ならば、商品Aを購入した客に対して商品Bの広告を送った場合）に期待できる利益増幅の大きさが相関ルールの評価値となる。たとえば、「商品A→商品B」という相関ルールでは、利益-コストが評価値となる。利益は（商品Bの売値-商品Bの仕入値）、コストは100と仮定しているので、評価値=（商品Bの売値-商品Bの仕入値）-100となる。

【0039】数は売れないが、1つ売れると多大な利潤の得られるような商品Dがあったとする。データマイニングの結果、そのような商品が右辺に来る相関ルールとして、「商品C→商品D」（商品Cを買う客は商品Dも買う）という相関ルールが抽出されたとする。この商品Dは、前述の通り、数は売れないものなので、当然、この相関ルールの支持度は低い。さらに、この相関ルールの確信度も低かった場合、この相関ルールを確信度と支持度で評価すると、価値の低い相関ルールということになる。しかし、この例のような評価尺度を用いた場合、1つ売れた場合の利潤が大きいために、支持度、確信度が共に低くても、高い価値が算出される場合がある。このように、確信度、支持度によって画一的に評価してしまうと他の相関ルールに埋もれてしまうような有効相関ルールを、この方法によって洗い出すことができる。

【0040】たとえば、新しい評価基準「利益の見込み」を設け、もしこの基準が大きな相関ルールがあり、その相関ルールの支持度、確信度が低いと仮定すると、相関ルール生成手段12は支持度、確信度のしきい値を小さくすればする程、膨大な数の相関ルールを抽出することになる。もし、「利益の見込み」という評価基準がなければ、ユーザはどの相関ルールを適用すればよいのか（大きな利益が見込めるか）、その選択に困ることになる。そこでこの評価基準の値を全ての相関ルールについて計算し、ある一定のしきい値以上の相関ルールを表示したり、評価基準の順に表示したりすることによって大きな利益の見込める相関ルールがユーザの目に止り易くなる。

【0041】

【発明の効果】この発明におけるデータマイニング装置は、評価尺度指定手段によって、データマイニングを行うユーザの目的にあった相関ルールの評価尺度をユーザ自身が指定し、その評価尺度に基づいて相関ルール評価手段が相関ルールの価値を算出し、実行結果表示手段は、この相関ルールの価値によって、出力する相関ルールの限定や並び替えを行うので、ユーザは、データマイニングの結果得られた大量の相関ルールの中から、自分

の設定した評価尺度を満足する、すなわち自分の目的に適合した相関ルールのみを抽出することが可能となる効果がある。

【0042】また、本発明の第2の発明のデータマイニング装置における評価尺度指定手段は、ユーザが相関ルールを利用した場合に発生する損失値または利益値（以下、相関ルール適用時のコスト）と、その相関ルールが成立した場合に発生する利益値または損失値（以下、相関ルール成立時の利益）を指定することにより評価尺度の指定を行うように構成されるので、ユーザにとってのメリット（利益値）とデメリット（損失値）を分けて入力することができ、評価尺度の設定が容易になる効果がある。

【0043】また、本発明の第3の発明のデータマイニング装置における評価尺度指定手段において、相関ルール適用時のコストおよび相関ルール成立時の利益は、データ項目毎に各種項目について値が定義されたデータ辞書中に定義された項目、または定数値、またはその両方を項とする、四則演算式を用いて述されるように構成されるので、ユーザは複雑な評価尺度を設定することが可能となり、ユーザが希望する評価尺度を柔軟に設定できる効果がある。

【0044】た、本発明の第4の発明のデータマイニング装置における評価尺度指定手段は、指定した相関ルール適用時のコストおよび相関ルール成立時の利益に関する情報を記憶装置に保存し、後でこれを読み込んで利用する機能を持つように構成されるので、一度設定した評価尺度を別のデータをマイニングするときにも適用することができるようになり、ユーザの利便性が向上する効果がある。

【0045】また、本発明の第5の発明のデータマイニング装置における相関ルール評価手段は、前記評価尺度指定手段によって指定された相関ルール適用時のコストおよび相関ルール成立時の利益、前記データベース中の相関ルールの精度（正解率）を示す確信度、並びに前記データベース中でその相関ルールが成立しているレコード数の全レコード件数に対する割合を表す支持度に基づいて評価値を算出するように構成されるので、利益値および損失値の設定が容易となる効果がある。

【0046】また、本発明の第6の発明のデータマイニング装置における相関ルール評価手段は、相関ルール適用時のコストおよび相関ルール成立時の利益を記述した四則演算式を計算する際に、データ辞書中に定義された項目の値を参照するように構成されるので、評価尺度を、より一般的な形式で記述することができる。たとえば、売り値や仕入れ値が変動した場合、データ辞書の変更のみで再度相関ルールの評価を行うことができ、評価尺度に変更を加える必要がない。その結果、評価尺度の普遍性が向上する効果がある。

【0047】また、本発明の第7の発明のデータマイニ

ング装置における実行結果表示手段は、演算の結果得られた相関ルールのうち、相関ルール評価手段が算出した評価値がユーザによって指定された基準値以上である相関ルールのみを表示する機能を持つように構成されるので、ユーザが設定した評価尺度に照らし合わせて価値の低い相関ルールを除外できる効果がある。

【0048】また、本発明の第8の発明のデータマイニング装置における実行結果表示手段は、演算の結果得られた相関ルールを、相関ルール評価手段が算出した評価値に基づき、昇順または降順の並べ替え処理を行った上で表示する機能を持つように構成されるので、ユーザが設定した評価尺度に照らし合わせて価値の高い相関ルールから参照することができる効果がある。

【図面の簡単な説明】

【図1】 この発明の実施の形態におけるデータマイニ

ング装置の構成図である。

【図2】 この発明の実施の形態におけるデータマイニング装置の処理の流れを示す図である。

【図3】 この発明の実施の形態における評価尺度の指定を行う画面の図である。

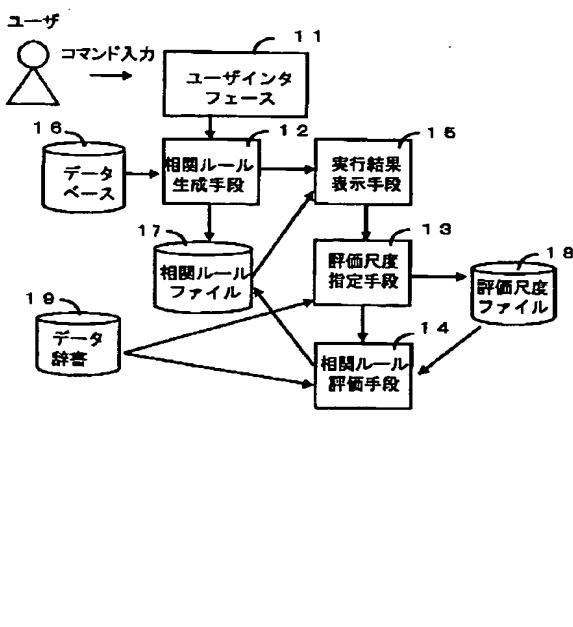
【図4】 この発明の実施の形態の具体例を説明するためのデータ辞書のイメージを示した図である。

【図5】 この発明の実施の形態における評価尺度の指定を行う画面の図である。

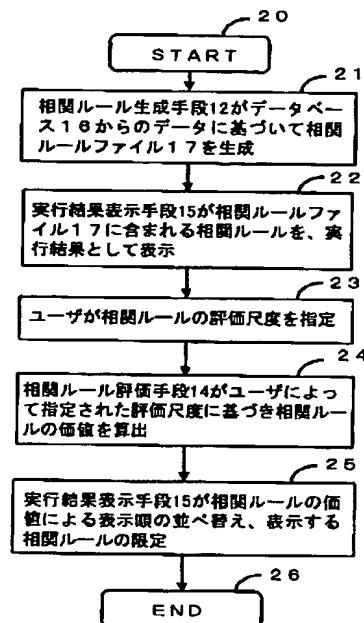
【符号の説明】

11 ユーザインターフェース、 12 相関ルール生成手段、 13 評価尺度指定手段、 14 相関ルール評価手段、 15 実行結果表示手段、 16 データベース、 17 相関ルールファイル、 18 評価尺度ファイル、 19 データ辞書

【図1】



【図2】



【図3】

評価尺度の指定			
評価尺度の名前 :	31		
相関ルール適用時のコスト :	32		
相関ルール成立時の利益 :	33		
34 右辺<項目名> 左辺<項目名>			
36 終了	37 中止	38 保存	39 読み込み

【図4】

データ辞書

	定価	売価	仕入値	...
パン	120	120	75	...
バター	360	330	290	
ミルク	130	120	80	
ジャム	250	225	205	
:	:	:	:	

【図5】

評価尺度の指定

評価尺度の名前 :  31

相関ルール適用時のコスト :  32

相関ルール成立時の利益 :  33

35

34 右辺<項目名> データ辞書内項目一覧  
 36  37  38 39 40  
 定価  
 売価  
 仕入値  
 値引率